

Developing a secure voice recognition service on Raspberry Pi

Van-Hoan Le¹, Nhu-Quynh Luc², Duc-Huy Quach²

¹AIoT System Joint Stock, Hanoi, Vietnam

²Academy of Cryptography Techniques, Hanoi, Vietnam

Article Info

Article history:

Received Oct 3, 2023

Revised Mar 21, 2024

Accepted Mar 28, 2024

Keywords:

Advanced encryption standard

Artificial neural network

Fast Fourier transform

Hidden Markov model

Raspberry Pi

Rivest–Shamir–Adleman

ABSTRACT

In this study, we present a novel voice recognition service developed on the Raspberry Pi 4 model B platform, leveraging the fast Fourier transform (FFT) for efficient speech-to-digital signal conversion. By integrating the hidden Markov model (HMM) and artificial neural network (ANN), our system accurately reconstructs speech input. We further fortify this service with dual-layer encryption using the Rivest–Shamir–Adleman (RSA) and advanced encryption standard (AES) methods, achieving encryption and decryption times well suited for real-time applications. Our results demonstrate the system's robustness and efficiency: speech processing within 1.2 to 1.9 seconds, RSA 2048-bit encryption in 2 to 6 milliseconds, RSA decryption in 6 to 10 milliseconds, and AES-GCM 256-bit encryption and decryption in approximately 2.6 to 3 seconds.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nhu-Quynh Luc

Academy of Cryptography Techniques

141 Chien Thang Road, Tan Trieu, Thanh Tri, Hanoi, Vietnam

Email: quynhln@actvn.edu.vn

1. INTRODUCTION

In the realm of digital communication, voice recognition technology emerges as a cornerstone, revolutionizing how humans interact with machines and digital interfaces. With its integration into a myriad of applications, from smart home devices to accessibility tools, voice recognition stands at the forefront of enhancing user experiences [1], [2]. Despite its rapid evolution and integration into our daily routines, a critical examination reveals a notable gap in the literature: the amalgamation of advanced security protocols within voice recognition systems remains underexplored. This oversight not only undermines the privacy and security of user data but also poses significant risks, making the need for fortified voice recognition services more pronounced than ever. Addressing this paramount concern, this study proposes an innovative approach by amalgamating the reliability of the hidden Markov model (HMM) [3], [4] and the adaptability of artificial neural network (ANN) [5]–[7] with the robust security afforded by the Rivest–Shamir–Adleman (RSA) encryption standard [8], all seamlessly integrated into the Raspberry Pi 4 model B platform. This pioneering service is designed to not just recognize voice in real-time with remarkable accuracy but to also safeguard the integrity and confidentiality of voice data against burgeoning cyber threats, strictly adhering to the latest National Institute of Standards and Technology (NIST) security standards [9].

Central to our innovation is the implementation of a 2048-bit RSA encryption, a gold standard in cryptographic security, ensuring that voice data converted into text for recognition is protected under the impenetrable veil of the public key cryptography standard (PKCS) #1 standard version 2.1. This meticulous integration of cutting-edge technologies culminates in the development of voice-SEC, a secure voice recognition service that represents a significant leap forward in the domain. Voice-SEC marks an advancement from our previous research as documented in [10], further showcasing our team's commitment to evolving secure voice recognition solutions. Through rigorous experimentation and evaluation, this study

not only validates the effectiveness and efficiency of voice-SEC but also illuminates its potential to redefine secure voice recognition technology. As we navigate through the subsequent sections, this paper will unfold the comprehensive development process, the theoretical underpinnings, and the empirical evidence supporting the efficacy of voice-SEC. From the meticulous design choices to the nuanced challenges and breakthroughs, we aim to provide a thorough exposition that not only showcases the technological advancements achieved but also sparks further research and innovation in secure voice recognition.

2. RELATED WORKS

2.1. Speech conversion and processing

The landscape of speech processing encompasses a variety of methods tailored for practical applications [9], [11], including the utilization of audiovisual toolboxes [12], group delay techniques [13], and notably, the fast Fourier transform (FFT) [14]. Among these, FFT stands out for its exceptional speed and efficiency in handling speech signals. Originating from the discrete Fourier transform (DFT), FFT addresses the inherent challenges of processing large speech samples by significantly reducing the computational complexity of FFT $\frac{N}{2} \log_2 N$. While FFT may exhibit a slight decrease in output information reliability when compared to DFT, its advantages in processing speed are undeniable, offering real-time performance capabilities essential for modern applications.

Opting for FFT as our primary method for speech conversion within this service was a strategic choice driven by its proven efficiency and speed. The FFT algorithm transforms speech into a digital format with remarkable speed, ensuring our service can operate in real-time without delay. The resulting output, a detailed spectrum of phonemes, is then processed using the HMM and ANN, providing a robust framework for accurate speech recognition. This integrated approach, combining FFT's rapid signal processing with HMM and ANN's sophisticated recognition capabilities, enables our service to deliver superior performance, making it a valuable tool for various digital communication needs.

2.2. Security solutions for voice recognition service

Expanding on the theoretical foundations detailed in publications [15]–[20], our team meticulously tailored and enhanced the technical facets of RSA encryption to suit the unique requirements of our application. This customization involved a comprehensive reengineering of RSA-based voice encryption techniques to ensure optimal compatibility and security, thus fortifying our system against a spectrum of cyber threats. In this critical phase of our research, we embarked on a quest to identify robust encryption solutions that could offer impenetrable security for digital data amidst the evolving landscape of cyber attacks. Our analysis juxtaposed the merits of two preeminent security measures: RSA 2048-bit [21]–[23] and AES-256 in GCM mode [24], [25]. Adhering strictly to the latest NIST guidelines, we ensured that our choice of encryption standards would not only meet but exceed the stringent requirements for digital voice signal protection, irrespective of the encryption paradigm employed. This meticulous approach to enhancing the security of the voice conversion process is depicted in Figure 1, providing a visual testament to our commitment to safeguarding digital communications against unauthorized access and ensuring the integrity of user data.

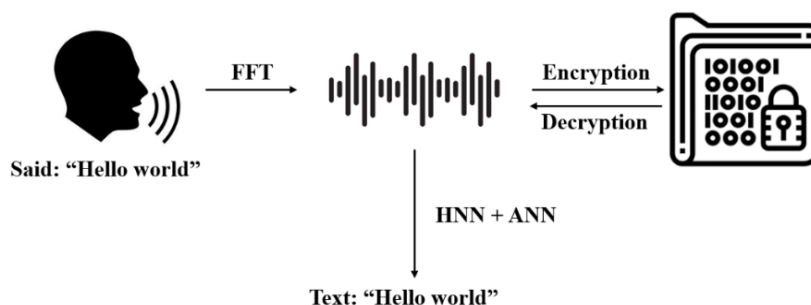


Figure 1. Voice information protection process

3. RESULTS AND DISCUSSION

3.1. Secure voice recognition service model

The secure voice recognition service is built on the Raspberry Pi 4 model B device. To access the service, users need to connect to the device through a Wi-Fi access point and then access the system's IP

address (192.168.1.10) to utilize the service (Figure 2). Each user will have a separate account to access the service. Upon successful login, the key associated with the account will be retrieved for use in speech encryption and decryption. The ciphertext of each user can only be decrypted by that specific user. The specific operational flowchart of the service is presented in Figure 3.

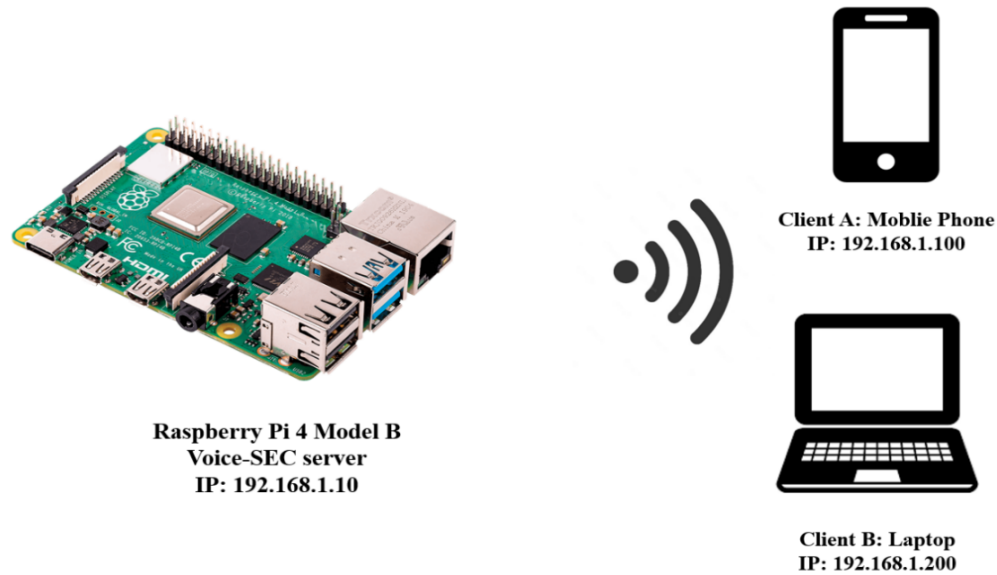


Figure 2. The operational model of the voice-SEC service

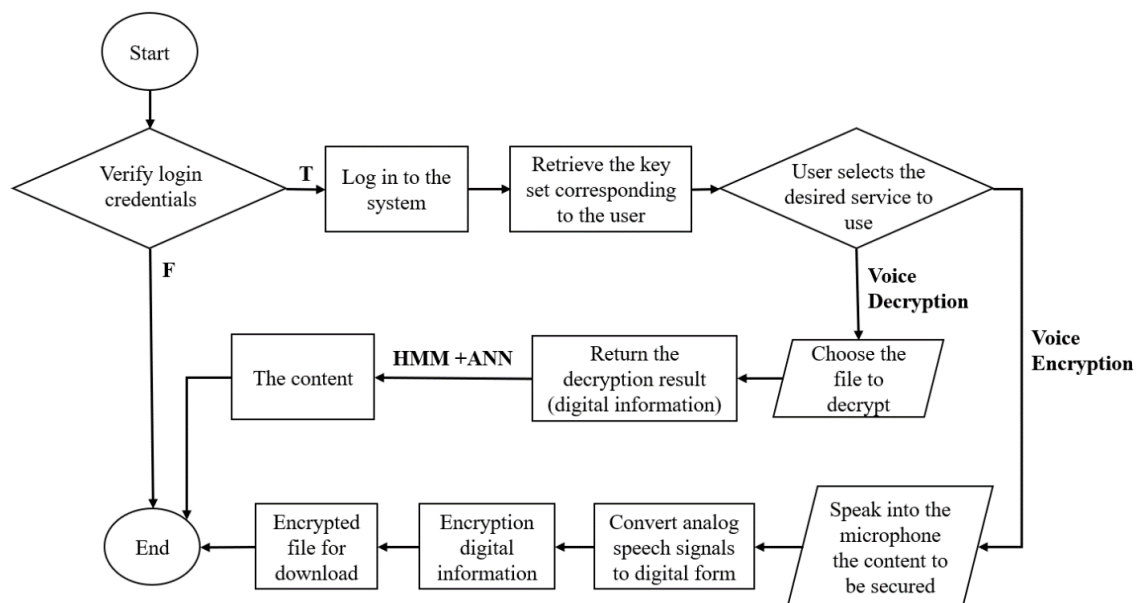


Figure 3. Flowchart of the voice-SEC service

3.2. Interface design for voice-SEC service

After accessing the device's IP address, a login interface like Figure 4(a) will appear. Users need to use the pre-established account credentials to log in and access the service. Upon successful login, users proceed to select the desired service using the left-side slider as shown in Figure 4(b).

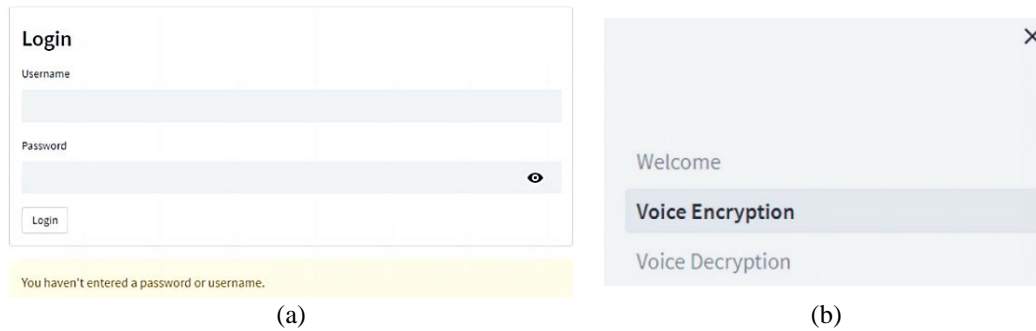


Figure 4. The login and usage interface of the system: (a) login interface of the service and (b) service selection bar

If the user selects the voice encryption service, they will be directed to an interface similar to Figure 5(a). Here, the user clicks on the microphone icon to start recording the content to be secured. After successful recording and processing, a download button will appear as shown in Figure 5(b). By clicking on this button, the user can download the encrypted file to their device. The encrypted file will have a structure similar to Figure 5(c).

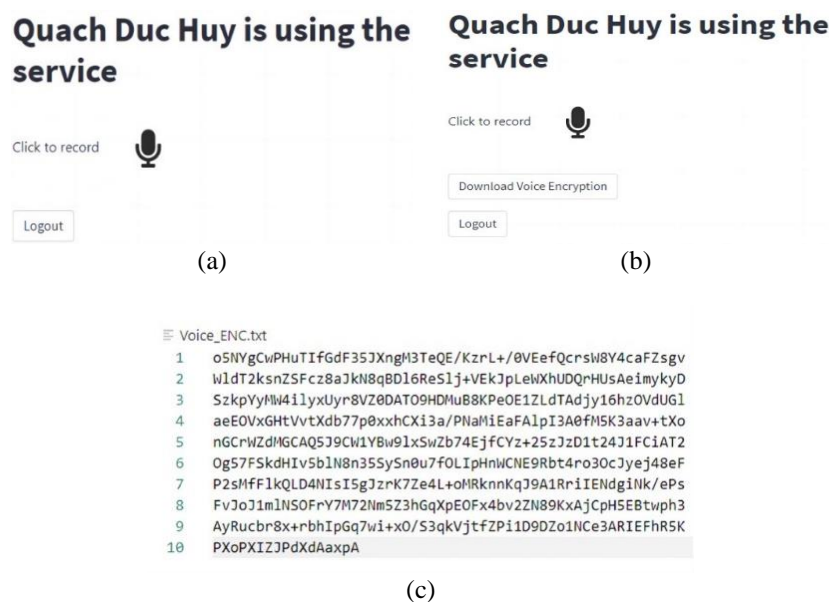


Figure 5. Voice encryption service interface: (a) voice encryption interface; (b) successful voice encryption interface; and (c) structure of the encrypted file

If the user chooses the voice decryption service, they will access the interface as shown in Figure 6(a). Here, the user uploads the encrypted file obtained from the encryption process. After successful decryption, the original message will be displayed on the screen as shown in Figure 6(b), and this content will be read out with a standard sample voice.

3.3. Evaluation of service deployment results on Raspberry Pi 4

In this study, to evaluate the execution efficiency of the Raspberry Pi 4 model B device, the author conducted experiments with various inputs for the voice-SEC service. Table 1 presents the results of running the voice-SEC service with different input data, where the key sets used all meet the quality evaluation standards set by the NIST. The results show that the execution time of the voice-SEC service using RSA 2048-bit encryption is approximately 1.2-1.9 s for speech processing, 2-6 ms for RSA 2048-bit encryption, and 6-10 ms for RSA 2048-bit decryption. The execution time of the voice-SEC service using AES-GCM 256-bit encryption is approximately 1.7-2 s for speech processing and 2.6-3 s for AES-GCM 256-bit

encryption and decryption. Both modes demonstrate efficient processing and fast encryption and decryption times, suitable for real-time user needs. The speech decryption process is also within an acceptable range (1-2 s).

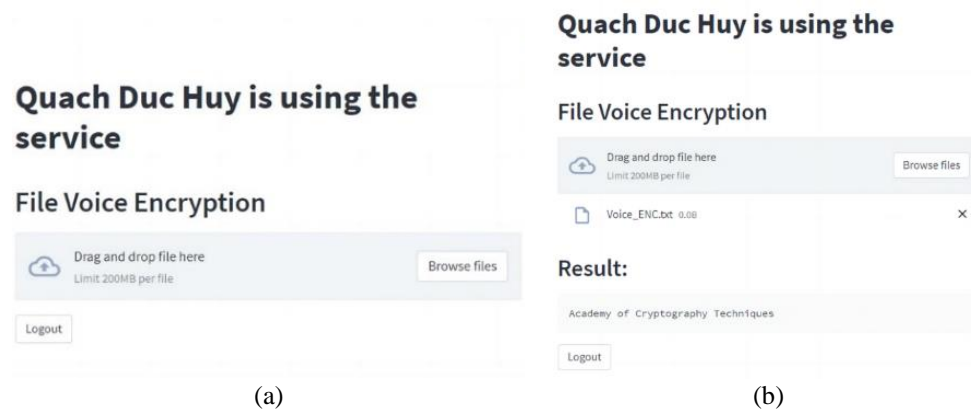


Figure 6. Voice decryption service interface: (a) voice decryption interface and (b) successful decryption interface

Based on the observed results (Table 1), the research team has found that the processing speed is directly proportional to the input length. However, there are instances where increasing the input length does not lead to an increase in processing time. This can be attributed to the fact that although the number of words in the input may be the same, the number of phonemes in a sentence may vary (some word clusters can be recognized as a single unit, while others need to be analyzed separately). Therefore, with the same input length, the processing speed of the service can still fluctuate.

Table 1. Results of running the voice-SEC service

Number of characters	Voice processing time (ms)	Encryption time (ms)	Decryption time (ms)
Voice security with RSA 2048-bit encryption			
12	1263.2608	3.991	7.0148
35	1379.675	1.9937	6.9827
37	1683.639	4.9993	5.9748
40	1830.3528	6.0288	9.9833
Voice security with AES-GCM-256-bit encryption			
21	1690.7975	2604.5635	2607.5628
27	1866.0744	2496.3576	2499.3501
35	1935.6286	3001.98	3004.9542

When comparing the performance of encryption and decryption between RSA 2048-bit and AES-GCM 256-bit encryption schemes, it is observed that RSA has a significantly faster encryption and decryption speed compared to AES-GCM 256-bit. However, when the length of the input speech content exceeds the length of the 2048-bit key, the encryption and decryption process in RSA may encounter errors, which is not the case with AES-GCM 256-bit encryption. Nevertheless, the authors prioritize the use of RSA 2048-bit as the primary security solution for the service because it can handle short input lengths by dividing the speech content into multiple blocks of 2048 bits for processing. For excessively large content blocks, the authors then opt for AES-GCM 256-bit as the security solution for the service.

As the research was conducted on a device, the research team also considered the packet transmission process between the device and the clients. After using Wireshark to capture packets during the transmission process, the research team obtained the results as shown in Figure 7. By establishing a secure certificate with RSA 4096-bit combined with SHA256 as shown in Figure 8, all packets are transmitted using the TLSv1.3 protocol combined with TCP, ensuring secure transmission of information between the device and the clients, where the transmitted information cannot be easily intercepted or understood by third parties.

The authors employed the Fortify static code analyzer toolkit (version 22.1.0.0166) to examine and evaluate the source code of the voice-SEC. Thorough findings obtained during testing, assessment, and

analysis of the source code of the voice-SEC program are presented in Table 2. The findings demonstrate that the voice-SEC has been developed without any flaws.

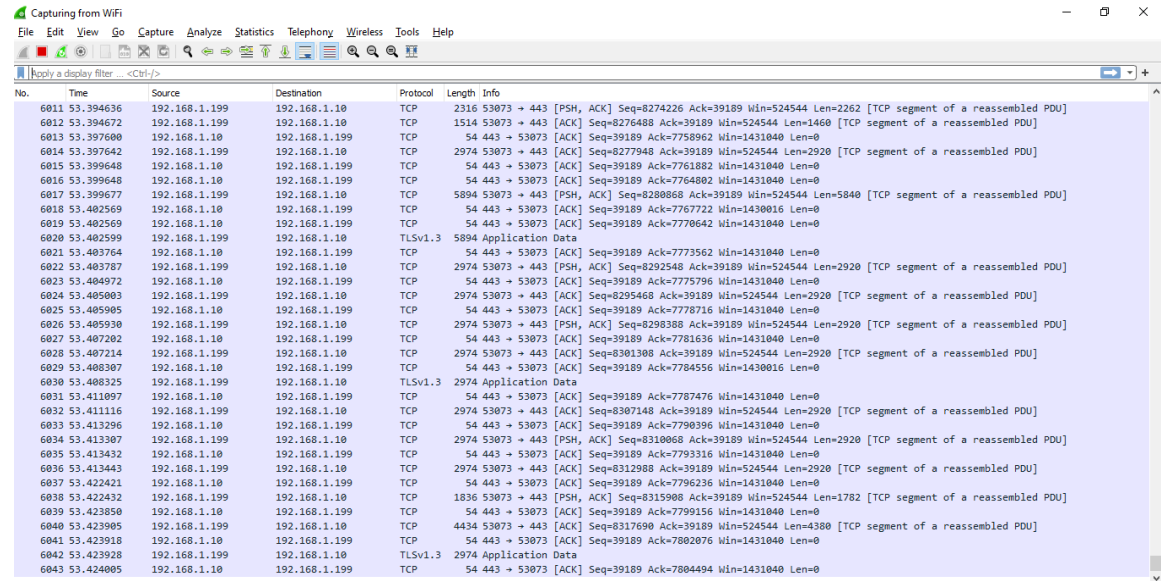


Figure 7. Packet capture results between the client and the device

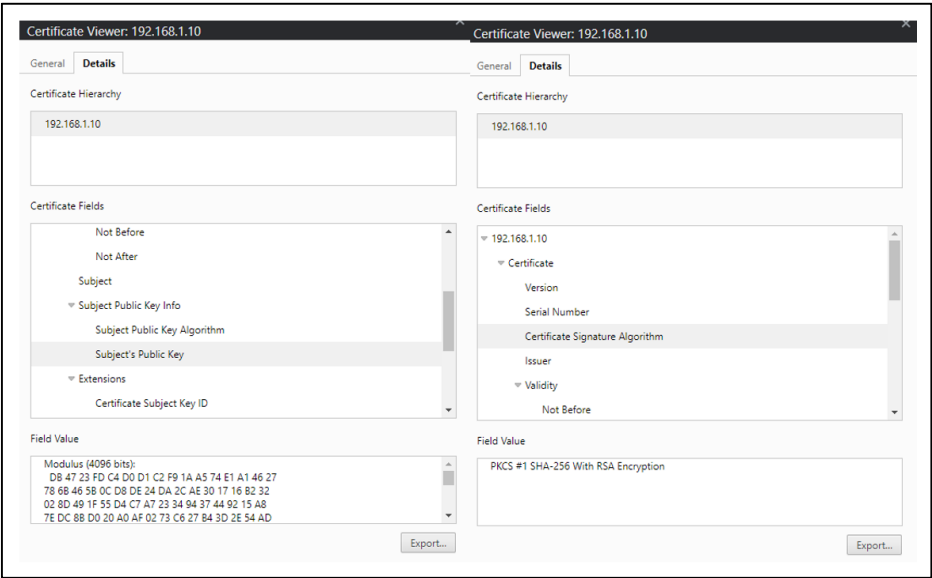


Figure 8. Certificate information used for securing the device

Table 2. Voice-SEC source code testing results using the Fortify static code analyzer

Category	Fortify priority (audited/total)				Total issues
	Critical	High	Medium	Low	
Buffer overflow	0	0	0	0	0
Poor style: variable never used	0	0	0	0	0
Type mismatch: integer to character	0	0	0	0	0
Type mismatch: signed to unsigned	0	0	0	0	0
Unchecked return value	0	0	0	0	0
Weak cryptographic hash	0	0	0	0	0

Thus, by analyzing, assessing, and testing the voice-SEC for flaws using the Fortify static code analyzer toolkit (version 22.1.0.0166). It is evident that the authors have developed and constructed the

software to ensure the safety of the source code. This provides sufficient evidence that the soft voice-RSA software can offer security against some of today's attacks when applied to real-world commercial applications that people use every day.

4. CONCLUSION

The accomplishments of this research mark a significant stride in the domain of secure voice recognition, thanks to the adept integration of the HMM and ANN with advanced encryption techniques like RSA public key cryptography (aligned with the PKCS#1 version 2.1 standard) and AES-GCM-256-bits. This dual-layer encryption framework not only ensures the utmost confidentiality of voice data but also demonstrates exceptional efficiency, with encryption and decryption times impressively ranging between 2-10 ms for RSA 2048-bits and 2.6-3 s for AES-GCM-256-bits, respectively. Despite these achievements, the study acknowledges certain limitations, notably in processing extensive speech inputs, which slightly curtail the system's applicability in more demanding scenarios. This recognition catalyzes future research endeavors aimed at transcending these boundaries. The forthcoming studies will delve into enhancing the system's capacity to handle larger datasets and exploring innovative encryption methodologies to further bolster security measures without compromising operational efficiency. Through continuous exploration and refinement, the research team is committed to advancing the field of secure voice recognition, paving the way for more resilient and versatile applications in the digital age.

ACKNOWLEDGEMENTS

The authors thank the Academy of Cryptography Techniques for supporting this work.





REFERENCES

- [1] N. Das, S. Chakraborty, J. Chaki, N. Padhy, and N. Dey, "Fundamentals, present and future perspectives of speech enhancement," *International Journal of Speech Technology*, vol. 24, no. 4, pp. 883–901, Dec. 2021, doi: 10.1007/s10772-020-09674-2.
- [2] X. Han *et al.*, "Pre-trained models: Past, present and future," *AI Open*, vol. 2, pp. 225–250, 2021, doi: 10.1016/j.aiopen.2021.08.002.
- [3] G. A. Fink, *Markov models for pattern recognition: From theory to applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, doi: 10.1007/978-3-540-71770-6.
- [4] Z. Han, Q. He, and M. von Davier, "Predictive feature generation and selection using process data from PISA interactive problem-solving items: An application of random forests," *Frontiers in Psychology*, vol. 10, Nov. 2019, doi: 10.3389/fpsyg.2019.02461.
- [5] I. Farkaš, P. Masulli, and S. Wermter, Eds., *Artificial Neural Networks and Machine Learning – ICANN 2020*, vol. 12397. Cham: Springer International Publishing, 2020, doi: 10.1007/978-3-030-61616-8.
- [6] G. R. Yang and X.-J. Wang, "Artificial neural networks for neuroscientists: a primer," *Neuron*, vol. 109, no. 4, p. 739, Feb. 2021, doi: 10.1016/j.neuron.2021.01.022.
- [7] R. Dastres and M. Soori, "Artificial neural network systems," *International Journal of Imaging and Robotics (IJIR)*, vol. 21, no. 2, pp. 13–25, 2021.
- [8] N. Bansal and S. Singh, "RSA encryption and decryption system," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 6, no. 5, pp. 109–113, Sep. 2020, doi: 10.32628/CSEIT206520.
- [9] E. Barker, "Guideline for using cryptographic standards in the federal government: Cryptographic mechanisms," Gaithersburg, MD, Mar. 2020, doi: 10.6028/NIST.SP.800-175Br1.
- [10] N.-Q. Luc, D.-H. Quach, C.-H. Vu, H.-T. Nguyen, and T.-L. Vo-Khac, "Integration of an RSA-2048-bit public key cryptography solution in the development of secure voice recognition processing applications," *Ministry of Science and Technology, Vietnam*, vol. 65, no. 3, pp. 3–7, Sep. 2023, doi: 10.31276/VJSTE.65(3).03-07.
- [11] E. K. Zaineb, S. Sahar, and M. Zouhir, "Pricing American put option using RBF-NN: New simulation of Black-Scholes," *Moroccan Journal of Pure and Applied Analysis*, vol. 8, no. 1, pp. 78–91, Jan. 2022, doi: 10.2478/mjpaa-2022-0007.
- [12] F. Ernawan, N. A. Abu, and N. Suryana, "Spectrum analysis of speech recognition via discrete Tchebichef transform," in *International Conference on Graphic and Image Processing (ICGIP 2011)*, Oct. 2011, vol. 8285, pp. 1619–1626, doi: 10.1117/12.913491.
- [13] S. Sadhu and H. Hermansky, "Radically old way of computing spectra: Applications in end-to-end ASR," in *Interspeech 2021*, Aug. 2021, pp. 1424–1428, doi: 10.21437/Interspeech.2021-643.
- [14] A. Abel and A. Hussain, "Multi-modal speech processing methods: an overview and future research directions using a MATLAB based audio-visual toolbox," in *Multimodal Signals: Cognitive and Algorithmic Issues*, A. Esposito, A. Hussain, M. Marinaro, and R. Martone, Eds. 2009, pp. 121–129, doi: 10.1007/978-3-642-00525-1_12.
- [15] S. F. Yousif, "Encryption and decryption of audio signal based on RSA algorithm," *International Journal of Engineering Technologies and Management Research*, vol. 5, no. 7, pp. 57–64, Mar. 2020, doi: 10.29121/ijetmr.v5.i7.2018.259.
- [16] E. Abouelkheir and S. El-Sherbiny, "Enhancement of speech encryption/decryption process using RSA algorithm variants," *Human-centric Computing and Information Sciences*, vol. 12, Feb. 2022, doi: 10.22967/HGIS.2022.12.006.
- [17] P. Pal, B. C. Sahana, S. Ghosh, J. Poray, and A. K. Mallick, "Voice password-based secured communication using RSA and ElGamal algorithm," in *Progress in Advanced Computing and Intelligent Engineering*, C. R. Panigrahi, B. Pati, B. K. Pattanayak, S. Amic, and K.-C. Li, Eds., Singapore: Springer Singapore, 2021, pp. 387–399, doi: 10.1007/978-981-33-4299-6_32.
- [18] M. M. Rahman, T. K. Saha, and M. A. Bhuiyan, "Implementation of RSA algorithm for speech data encryption and decryption," *IJCSNS International Journal of Computer Science and Network Security*, vol. 4, no. 3, pp. 337–351, 2012.





- [19] S. Al-Ghamdi and H. Al-Sharari, "Improve the security for voice cryptography in the RSA algorithm," in *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, IEEE, Feb. 2022, pp. 1–4, doi: 10.1109/ICBATS54253.2022.9759016.
- [20] C.-L. Duta, L. Gheorghe, and N. Tapus, "Real-time DSP implementations of voice encryption algorithms," in *Proceedings of the 3rd International Conference on Information Systems Security and Privacy*, SCITEPRESS - Science and Technology Publications, 2017, pp. 439–446, doi: 10.5220/0006208304390446.
- [21] Z. Zheng and F. Liu, "On the high dimensional RSA algorithm—A public key cryptosystem based on lattice and algebraic number theory," *arXiv*, doi: 10.48550/arXiv.2202.02675.
- [22] L. Matysiak, "Generalized RSA cipher and Diffie-Hellman protocol," *Journal of Applied Mathematics and Informatics*, vol. 39, no. 1–2, pp. 93–103, 2021, doi: 10.14317/jami.2021.093.
- [23] N. Stoianov and A. Ivanov, "Public key generation principles impact cybersecurity," *Information & Security: An International Journal*, vol. 47, no. 2, pp. 249–260, 2020, doi: 10.11610/isi.4717.
- [24] N. Ahmad, L. M. Wei, and M. H. Jabbar, "Advanced encryption standard with Galois counter mode using field programmable gate array," *Journal of Physics: Conference Series*, vol. 1019, p. 012008, Jun. 2018, doi: 10.1088/1742-6596/1019/1/012008.
- [25] S. M. Soliman, B. Magdy and M. A. Abd El Ghany, "Efficient implementation of the AES algorithm for security applications," *2016 29th IEEE International System-on-Chip Conference (SOCC)*, Seattle, WA, USA, 2016, pp. 206–210, doi: 10.1109/SOCC.2016.7905466.

BIOGRAPHIES OF AUTHORS







Van-Hoan Le     has a Master's degree in Electronics and Communication, and he is currently managing a joint-stock company named AIoT, specializing in electronic device systems. He can be contacted at email: hoanle.aiot@gmail.com.



Nhu-Quynh Luc     was born in Nam Dinh, Vietnam in 1983. He received his Bachelor's in Mathematics at the Vietnam University of Science (VNU) in 2006, and a Master of Crypto-graphic Engineering at the Academy of Cryptography Techniques, Vietnam. He has a Ph.D. in Electronic Materials Science. In his B.E. and Master's, he focused on the mathematical aspects of elliptic curves and their applications in cryptography. Currently, his research interests are organic materials and their use in micro-electronic fabrication. He can be contacted at email: quynhln@actvn.edu.vn and lucnhuquynh69@gmail.com.



Duc-Huy Quach     was born in Vietnam in 2000. He studied at the Academy of Cryptography Techniques, in Vietnam. He focused on the mathematical aspects of post-quantum cryptography and its applications in cryptography. He can be contacted at email: qdhuy2000gl@gmail.com.